Sex-specific impact of inbreeding on pathogen load in the striped dolphin

Georgios A Gkafas[1,2], Menno de Jong[1], Athanasios Exadactylos[2], Juan Antonio Raga[3], Francisco J Aznar[3] & A Rus Hoelzel[1,§]

[1]Department of Biosciences, Durham University, DH1 3LE, South Road, Durham, UK

[2]Department of Ichthyology and Aquatic Environment, School of Agricultural Sciences, University of Thessaly, Fytokou str. 384 46, Volos, Hellas

[3]Unidad de Zoología Marina Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universitat de València, Aptdo 22085, E-46071-Valencia, Spain

[§] Corresponding Author: A. Rus Hoelzel. School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, UK. Tel: +44 191 334 1325; fax: +44 191 334 1201; email: a.r.hoelzel@dur.ac.uk

Abstract

The impact of inbreeding on fitness has been widely studied and provides consequential inference about adaptive potential and the impact on survival for reduced and fragmented natural populations. Correlations between heterozygosity and fitness are common in the literature, but they rarely inform about the likely mechanisms. Here we investigate a pathology with clear impact on health in striped dolphin hosts (a nematode infection that compromises lung function). Dolphins varied with respect to their parasite burden of this highly pathogenic lung nematode (*Skrjabinalius guevarai*). Genetic diversity revealed by high resolution restriction associated DNA (43,018 RADseq SNPs) analyses showed a clear association between heterozygosity and pathogen load, but only for female dolphins, for which the more heterozygous individuals had lower *Skrjabinalius guevarai* burden. One locus identified by RADseq was a strong outlier in association with parasite load (heterozygous in all uninfected females, homozygous for 94% of infected females), found in an intron of the Citron Rho-Interacting Serine/Threonine Kinase (CIT) locus (associated with milk production in mammals). Allelic variation at the Class II MHC DQB locus was also assessed and found to be associated with both regional variation and with pathogen load. Both sex specificity and the identification of associating functional loci provide insight into the mechanisms by which more inbred individuals may be more susceptible to the infection of this parasite. This provides important insight towards our understanding of the impact of inbreeding in natural populations, relevant to both evolutionary and practical conservation considerations.

1. Introduction

Genetic diversity affects both short-term individual fitness and long-term population adaptive potential, and these factors are inter-dependent. Populations need to retain diversity to respond to new selection pressures in a changing environment, including pathogen challenges, which may in turn be affected by environmental change. A relationship between diversity and fitness has been shown in small populations where susceptibility to pathogens can be promoted by the loss of heterozygosity due to both inbreeding and genetic drift (e.g. [1]). In fact this relationship holds for a broad range of population sizes, and there have been many studies reporting heterozygosity-fitness correlations (HFCs; see reviews in [2, 3]) and assessments of the most effective measures of inbreeding to identify these correlations (see [3]). Many earlier studies used microsatellite DNA markers, but more recent studies demonstrate that the greater power availed by genome sampling reveals patterns that may otherwise have been missed (e.g. [4, 5, 6]). The restriction associated DNA (ddRADseq; [7]) method employed here screens across the genome at thousands of loci.

A positive correlation could be associated with a particular locus, loci in linkage disequilibrium (LD) with the marker loci, or it may reflect a more general pattern of inbreeding across the genome (see [8]). In a study on Galapagos sea lions (*Zalophus wollebaeki*), the relative importance of genome-wide effects compared to specific loci (in that case the major histocompatability complex 'MHC' loci) were investigated [9]. The authors controlled for genome-wide inbreeding to test the influence of the MHC DRB locus and found strong associations between DRB diversity and all tested fitness traits (which included birth mass, pup survival and female reproductive success), indicating that single-locus effects can be important.

78 At the same time, from a review of the literature, Szulkin et al. [10] concluded that
79 HFC can most often be explained by inbreeding, which affects the whole genome.
80 Evidence for balancing selection at MHC loci to retain diversity is extensive,
81 and defense against pathogens is the likely driver (e.g. [11]). This is known to be an
82 important factor in marine species, including cetaceans [12, 13]. Diversity at these
83 loci may also be promoted by sexual selection, based on evidence that females in
84 some species select mates based on their MHC genotype [14, 15]. Vassilakos et al.
85 [13] found evidence for regional variation in the charge structure of the peptide
86 binding region of the MHC DQB locus (among the 'pocket 4' residues that
87 influence T-cell recognition; after [16]) for two species of cetacean (*Orcinus orca*
88 and *Tursiops truncatus*), which may imply fitness variation among populations
89 associated with local adaptation.
90 Here we study inbreeding in striped dolphins (*Stenella coeruleoalba*) with
91 variable burdens of the parasitic lungworm, *Skrjabinalius guevarai*. Lungworm
92 nematodes are common in the delphinid respiratory system [17], and this species of
93 lungworm can cause almost total occlusion of bronchi and bronchioles [18, 19]),
94 thereby imposing a significant impact on health. We test the hypothesis that there
95 will be a correlation between genome-wide measures of heterozygosity and
96 pathogen load, but also look for strong effects at single loci (from a genome scan),
97 and at exon 2 of the MHC DQB locus, including pocket 4 charge properties in the
98 peptide binding region (see [13]). Given the potential for a differential impact of
99 parasite load on males and females (e.g. associated with the added energetic
100 requirements of females during pregnancy), we also test the hypothesis that the
101 relationship between parasite load and inbreeding will differ between males and
102 females.

2. Materials and Methods

(a) Samples and sex identification

Eighty four striped dolphins from the western Mediterranean Sea (near Valencia, Spain) were collected as stranded animals between1990 & 2008 (see Table S1). Of these 51 were collected during periods of morbillivirus epizootics [20] and 33 were outside those time periods. Morbillivirus infection was not confirmed for the samples collected during the epizootics and used in this study. Dolphins were transferred to the laboratory where necropsy and anatomical analyses were carried out immediately (following the protocols of [21]), or alternatively stored at -20^{o}C for later analysis. Sex was identified visually or by PCR on DNA extracted using standard protocols (see [22]) using the primers P15EZ, P23EZ for the Zfx/Zfy gene (after [23]) or Y53-3c and Y53-3d for the SRY gene [24]. We considered dolphins longer than 160 cm to be at least several years old (after [25]). There were 8 out of 68 for the ddRAD analyses and 16/ 80 for the MHC analyses that were shorter than 160 cm. In case parasite load is associated with age (e.g. due to accumulation with time or the nature of infection), we repeated all analyses excluding the younger dolphins. All results that had been significant for the full dataset remained significant (data not shown).

(b) Parasite analysis

Lungs were removed for parasite analysis and each lung was weighed to the closest milligram. The lung was always opened starting from the main bronchus of the upper lobe which is connected to the trachea, and then the duct of each bronchioles and alveoli were followed through to the end of the bottom lobe (see Figure S1). Only whole parasites or the parasites' tail were collected, and then

stored in saline buffer to maintain a constant pH and isotonic environment [26]. After cleaning with the isotonic buffer, parasites from each lung were preserved in 70% alcohol. After gross examination, lungs were washed out on a 0.2 mm sifter and any parasites (whole or tails) obtained were collected. All parasites were examined in a stereoscope for species identification. Furthermore, 10% of the total number of parasites were prepared and screened under a microscope to ensure the consistency of species identification. A Petri dish with divided areas was used for the parasite counting. Parasites of each lung were combined for the total individual lung-parasite burden.

There was a relatively low number of worms (1 – 20; for both lungs combined) for some infected individuals, and visual examination indicated that this number of worms was not sufficient to occlude the bronchioles or alveoli. At the same time, a medium or high level of infestation resulted in obvious occlusions. This may suggest a threshold value above which an impact may begin to be seen. For this reason, analyses included the comparison of two infestation categories; none/low infection (0 to 20 parasites) and medium/high (>20 parasites) infection. However, it is possible that pathology is also or primarily related to secondary bacterial infections, though we have no data on this for these animals. For example, *Torynurus convolutus* and *Pseudalius inflexus* lung infestation led to secondary bacterial infections responsible for mortality of North and Baltic Sea harbour porpoise (*Phocoena phocoena*; [27])]. Therefore, we also consider the comparison between lungworm infected and uninfected dolphins.

(c) Genomic analyses

We constructed a DNA library of 84 samples (7 pools of 12 samples each) following the ddRADseq protocol described in [7]. We chose a 6 bp cutter (HindIII) and a 4 bp cutter (MspI) based on in silico simulations with the R package SimRAD [28]. The fragment size selection window was 250 - 350 bp with a size range of 100bp (selected using a Sage Science PippinPrep). Sequencing was paired end (2X 125bp) in one lane on an Illumina HiSeq_2500 (version 4 chemistry). Reads were trimmed to 110 bp and demultiplexed using the process_radtags command of the software STACKS [29]. After quality control (rejecting samples with less than 1 million reads) 68 samples were retained and there was an average of 2.8 million read pairs per sample among these. Paired reads were mapped against the *Tursiops truncatus* genome (accession GCA_001922835.1; [30]}) using BWA v. 0.7.12 (bwa mem –aM; [31]). Each resulting sam file was converted to bam format using SAMTOOLS v.1.3. [32].

Using the command SelectVariants, indels and non-biallelic SNPs were filtered out. Then using the command VariantFiltration, SNPs were filtered based on mapping quality using the following settings: --filterExpression "QD<2.0||FS>60.0||MQ<40.0||MQRankSum←12.5||ReadPosRankSum←8.0". The QUAL score (QD) was normalized by allele depth (AD) for a variant, and the Phred- scaled p-value (FS) used Fisher's exact tests to detect strand bias. The MQRankSum command set the Z-score from a Wilcoxon rank sum test of Alt vs. Ref read mapping qualities, and ReadPosRankSum did this for read position bias. SAMTOOLS was also used to pick up reads in concordance and retain SNPs with a single hit. Loci were assembled using the GATK HaplotypeCaller [33].

Using the -filterAlign plugin through TASSEL v.5.0 [34], the vcf file was filtered to require a minimum of 80% of taxa for which the SNP must have been

scored and a minor allele frequency (MAF) of 0.05. These settings generated 83,414 SNPs. Vcf-tools [35] through the command --thin (set at 200bp) was then used to retain a single SNP per read, reducing the final number to 43,018 SNPs for further analyses. The software TASSEL v.5.0 was also used for General Linear Model analysis. As required by TASSEL, the vcf file was sorted using the SortGenotypeFile plugin. For the Generalized Linear Model (GLM) analyses the filtered vcf file and the trait file (samples allocated to different infestation categories, age class and mortality during a morbillivirus event) were merged through the -intersect command. For the final GLM analysis, permutations were set to 1,000, under the *-FixedEffectLMPlugin* command. This function performs association analysis using a least squares fixed effects linear model and utilizes a fixed effects linear model to test for association between segregating sites and a trait. The qqman package [36] was used to visualize Manhattan plots and QQ plots of the outputs of the GLM analyses. The program plots the negative logarithm of p-value for each SNP across the genome. Bonferroni type one correction was used to assess significance for multiple tests.

The samples were considered to be from a single population, and to confirm this we used the PCA method implemented in the R package Adegenet version 2.0 [37]. To test for presence of SNPs on sex chromosomes we used the Perl script nucmer in the program MUMmer [38] to align the *Tursiops* reference genome against the Cow genome (accession number: GCA_002263795.2; [30]). The program finds maximal exact matches and aligns them to join the clusters into a single high scoring pair-wise alignment. The 'delta' file generated by this analysis was filtered using the --delta-filter flag. The program show-coords was then used to

parse the delta alignment output displaying summary information such as position, percent identity and percent alignment coverage.

We compared pathogen loads against several metrics of genomic diversity derived using the R package InbreedR [39]. The first is standardised multilocus heterozygosity (*sMLH*; see [10]), which assesses average heterozygosity across the genome. We used the second metric, g^2 [40], as a proxy for identity disequilibrium, providing an estimate of variation in identity by descent (e.g. if g^2=0 there is no variance in inbreeding in the sample). We also show the results from an alternative method for assessing identity disequilibrium, heterozygosity-heterozygosity correlation coefficients (HHC; [41]). This analysis reiterates the comparison of random subsets to show the distribution of HHC in the sample. We provide this as an illustrative metric, since it is less robust as a statistic that g^2 given that samples within the HHC distribution are non-independent. To further consider the role of inbreeding, we ran two analyses (see [10, 42]), one for the expected correlation between the trait value (*W*, pathogen load in this case) and heterozygosity *(h)*: r^2_{Wh}, and one for the expected correlation between inbreeding level (*f*) and the trait value (r^2_{Wf}). All analyses quantifying diversity assessments for the SNP dataset were repeated for the 23 microsatellite DNA locus dataset, which was analysed for the same set of samples and reported in [22].

(d) MHC analyses

Exon-2 of the MHC Class II DQB1 locus was amplified with DQB1 F: CTGGTAGTTGTGTCTGCACAC & DBQ1 R: CATGTGCTACTTCACCTTCGG (after [43]). Reaction conditions were 10mM Tris-HCl, 50mM KCl, 2,5mM MgCl2, 0.2mM of each dNTP, 0.25µM of each primer, 2 units of high fidelity Pfu

Taq polymerase (Promega, UK), 0.8mM DMSO 20% and 1µl of total DNA in 20µl final volume. The PCR cycling profile was an initial denaturation step at 95°C for 15 minutes, following by 30 cycles of denaturation at 95°C for 1 minute, annealing at 55°C for 30 seconds, and elongation at 72°C for 30 seconds followed by a final elongation step at 72°C for 15 minutes. To identify allelic diversity individuals were screened by Single Strand Conformation Polymorphism (SSCP) analysis [27]. Allelic conformation was visualized by exposure to UV light. The allelic diversity for each individual was scored and genotypes were assigned. After the identification of putative unique alleles, the same PCR products were loaded again onto a non-denaturing acrylamide gel (6%) and this time the bands representing unique alleles (with some replication) were extracted from the gel.

Gel fragments were crushed in 50µl of 10mM T.E. and incubated overnight at 37°C. One µl of the solution was then amplified by PCR (using the same concentrations and PCR profile as described before) using the high fidelity Pfu Taq polymerase. PCR products, prior to sequencing, were purified using a Qiagen QIAquick PCR purification Kit™, to remove primer dimmers, unincorporated dNTPs and chemicals, according to manufacturer instructions. Purified DNA was sequenced in both directions on an ABI 377 automated sequencer. The PCR products of the putative unique alleles were cloned, using Easy T-Vector Cloning Kit (Promega) according to manufacturer instructions, in order to compare allele sizes and confirm that a single band represented a single allele. Up to 8 clones were screened by SSCP from different individuals. A total of 80 individuals were successfully genotyped from the Valencia population. For this analysis and additional 22 striped dolphin samples from Ireland were also genotyped to consider the possibility of regional variation at the MHC loci (as reported earlier [13]). The

samples from Ireland were from stranding events outside of morbillivirus periods. Sequences were analysed using ChromasPro v. 1.5 (Technolysium Ltd.). Nucleotide sequences were aligned using ClustalX v. 2.0.12 [44]. BLAST (http://www.blast.ncbi.nlm.nih.gov/Blast.cgi/) was used to confirm that DNA sequences represented the exon-2 MHC Class II DQB1 locus. Rates of non-synonymous and synonymous substitutions were calculated using the software MEGA v. 6 [45]. The d_N / d_S ratio was computed according to the Nei-Gojobori method [46].

Amino acid distributions were calculated in the 10 residues of the peptide binding region (PBS; after [47]) to test for associations with specific functional components and parasite load. The charge of amino acids of P4 pocket was based on the *β70 β71 β74* residues according to physicochemical properties [48]. The amino acids' supertype state was determined according to the following categorization [49]:

(n) Neutral supertype: *F, M, W, I V, L, A, P, C, N, Q, T, Y, S, G*

(+) Positive supertype: *H, K, R*

(-) Negative supertype: *D, E*

The total charge of each allele was the sum of each residue's charge. For example if an allele was positively and negatively charged in the Pocket 4 it was classified in the di-charged supertype group. A Generalised Linear Model (GLM) was performed to evaluate associations between MHC genotype and gender, length, morbillivirus epizootic event periods and parasite load. The morbillivirus epizootic events were categorised as: 1) samples from the first recorded event from1990 to 1992, 2) samples from 1993 to 2006 which were outside of the morbillivirus epizootic events, and 3) samples from after 2007 during the second

morbillivirus epizootic. Parasite load was used as the response variable. Statistics were conducted using the statistical package R-platform. False Discovery Rates were determined using the largeQvalue software package [50].

3. Results

(a) Parasites

Comparing dolphins collected during or outside the time of morbillivirus epizootics showed no significant differences in genetic diversity, and no clear distortion of the pattern observed for lungworm infection (for either SNP or MHC analyses; data not shown), so all samples were used for further assessment based on lungworm parasite load. For the SNP dataset, 42 animals were infected with lungworm (61.8%), whereas 26 (38.2%) were uninfected. Only a single parasite species was found during the gross lung examination, the nematode *Skrjabinalius guevarai* (Nematoda: Pseudaliidae). The infestation load among individuals ranged from 0 to 370 worms. Lungworm counts fit a negative binomial distribution (skewness measure=0.159, with respect to the negative binomial, $p<0.05$). Parasite count intensity parameters (skewness, mean, median, exact confidence intervals) are shown in table S2. There were 18 infected individuals out of the 30 female striped dolphins, and 24 infected out of 38 males (Table S1) and these ratios are not significantly different (Contigency Test: Pearson chi-square= 0.02, P = 0.8875).

(b) Genomic measures of diversity

PCA analyses revealed a single cluster with a few outliers (figure S2), so the sample set was treated as a single population. All results were essentially the same when those outliers were removed (see Supplementary file and below), and so all

samples were retained. No significance was found for any tested associations between parasite load and genomic diversity as estimated using 23 microsatellite DNA markers (e.g. figure S3). Further analyses reported are therefore based on the 43,018 ddRAD SNP dataset. From our mapping of the *Tursiops* genome against cow (*Bos Taurus*, sequenced to chromosomes), we identified contigs in the *Tursiops* genome that map to the cow X-chromosome (see table S3). None of our SNPs map to those contigs, and so our results will be unbiased by sex-linked loci. For the SNP data, standardized Multilocus Heterozygosity (*sMLH*) was strongly correlated with parasite load in female individuals, but not in males (based on all individuals; see results in figure 1). This remains true when only infected individuals are considered (females: $r^2_{Wh} = 0.517$, $F_{1,16} = 17.14$, $p = 0.0007$; males: $r^2_{Wh} = 0.030$, $F_{1,22} = 1.804$, $p = 0.193$; table 1), and when the three female samples with the lowest sMLH were removed ($r^2=0.26$, $F_{1,13}=8.08$, $P= 0.0089$).

Mean *sMLH* was significantly elevated in low level compared to high level infected female individuals (*sMLH*= 1.10 ± 0.046 (s.d.) vs 0.948 ± 0.177, respectively; Mann-Whitney U test, Z= 2.523, p = 0.0114; figure 2; see table 1 for variance values). For males there was no significant difference (0.966 ± 0.170 vs 0.961 ± 0.132; Z= 1.40, p= 0.132). A highly significant relationship was also obtained between parasite load and measures of genome-wide inbreeding in female individuals, but not in males (see r^2_{Wh} and r^2_{Wf} in table 1). Analysing all 68 samples together, g^2 (assessing inbreeding among loci; [10, 38]) was significantly different from zero (p = 0.001), implying inbreeding (table 1). The distribution of heterozygosity– heterozygosity (*het-het*) correlation coefficients (HHC) show tight distributions for both SNP (table 1, figure S3) and microsatellite DNA data (figure S4) with a mean close to one, suggesting inbreeding. The GLM analysis

(implemented in TASSEL) showed a highly significant association between parasite load and a specific SNP, but this was revealed only in female individuals ($p= 8.21 \times 10^{-11}$; figure S5) after correction for Type I error (threshold $p = 1.16 \times 10^{-6}$). This SNP is found within the intronic region of the Citron Rho-Interacting Serine/Threonine Kinase (CIT) gene that functions in cell division. Uninfected females were all heterozygous at this SNP, and all but one infected individuals were homozygous (table 2). GLM analyses based on age class and/or mortality during morbillivirus events in conjuction with infestation categories for each gender did not reveal any strong associations as illustrated in the QQ-plots (figure S6). All analyses were replicated omitting the four outlier individuals from the PCA analysis shown in Figure S1. There were no differences in the patterns observed or levels of significance, illustrated by a replication of Table 1 omitting these samples in Table S4.

(c) MHC variation

Sequenced MHC clones revealed no more than two sequences in each individual. Twenty one alleles were found with a unique amino acid composition (table S5). Alleles were named Sc-DQB*01 to Sc-DQB*21 according to their frequency in the study population. A Blast search indicated amplification of the correct locus. Only one individual was homozygous at this locus, and so a test between parasite load and heterozygosity was not possible. In pocket4 of the PBS region, 100% of the translated amino acids were variable. The rate of nonsynonymous (d_N) compared to synonymous substitutions (d_S) was significantly elevated at antigen binding sites and within the P4 region (table 3). There was a significant difference in the P4 charge property profile comparing the populations

in Valencia and Ireland (χ^2 = 9.16, d.f. = 2, p = 0.01; figure S7). In the Valencia population we compared parasite load levels with allelic diversity. We found two alleles (Sc-DQB*11 and Sc-DQB*21) that were disproportionately likely to be present in individuals with no parasites (binomial test for the presence of either or both alleles calculating the combined probability (frequency) of these alleles in dolphins with parasites, and then determining the probability of finding none among the dolphins without parasites, p = 7.02 x 10^{-8}; table S3). GLM analyses did not reveal any associations between factors listed in table S6.

4. Discussion

Our RADseq analyses showed that genome-wide heterozygosity was significantly associated with lungworm infection, especially beyond a stage of infection that reflects substantial blockage of airways, however only in females. This could have important implications both for understanding evolutionary process (e.g. if selection affected males and females separately as a consequence), and for developing conservation strategy (e.g. if females are more impacted by low effective population size (Ne) and the loss of diversity than males). Lungworm infection is widespread in marine mammals [51] impacting on the health of both cetaceans and pinnipeds (e.g. [4]). Therefore inference drawn from our study could have implications for a broad range of other species.

A significant result restricted to females may be due to maternal stress factors such as parturition or nursing, causing females to cross a threshold such that the association with parasite resistance becomes apparent. In a study of Seychelles warbler (*Acrocephalus sechellensis*) Richardson et al. [52] found that the offspring of highly heterozygous females survived better than the offspring of inbred

mothers, potentially associated with female condition. They found no HFC for males. Jamieson et al. [53] also showed that the mother's level of inbreeding affects offspring fitness in the New Zealand takahe (*Porphyrio hochstetteri*). For the song sparrow (*Melospiza melodia*) Keller [54] showed that a reduction in fitness associated with inbreeding was only seen in inbred female individuals. Although we have no pedigree data, and so cannot assess impact on the health of offspring, a direct relationship between maternal and offspring health is often implied.

Relevant to this, we found a female-specific effect at the Citron Rho-Interacting Serine/Threonine Kinase (CIT) locus, discovered from screening the RAD data across the genome. CIT functions in cell division. Together with kinesin (KIF14), this protein localizes to the central spindle and mid-body of the cell, and functions to promote efficient cytokinesis. Smith et al. [55] argue that a greater increase in the percentages of bi-nucleated/ multinucleated cells were seen after expression of EGFP-bSV-831–1281, which contains a coiled-coil sequence and binding sites for the central-spindle protein KIF14. CIT is required for KIF14 localization to the central spindle and mid-body, so there is connection between CIT and the presence of bi-nucleated/ multinucleated cells, since CIT plays a role in cytokinesis and displays the serine/ threonine protein kinase activity. Rios et al. [56] suggest that bi-nucleated cells evolved to maximize milk production and promote the survival of offspring across all mammalian species through the expression of serine/threonine kinases (AURKA and PLK-1) as key-regulators of cytokinesis. This may suggest a more direct connection with fitness, whereby maternal inbreeding depression could affect the survival of offspring due to processes associated with lactation.

However, it isn't clear why CIT heterozygotes in particular would be associated with reduced lungworm infection. Furthermore, the SNP appears in an intron, and therefore is not certain to affect the structure or expression of the CIT locus (though intronic mutations can affect gene expression with or without an impact on alternative splicing; e.g. [57]). Given that our scan was based on a finite number of SNPs and referenced against a related species (*Tursiops truncatus*), it is likely that there are other relevant loci not identified by our analysis, and possible that the SNP identified is actually in LD with another locus that reflects the true function affected. Furthermore, there could be a heterosis effect at CIT or some other locus, such that the relevant function is improved for heterozygote females, but further work would be required to assess this and better understand the mechanisms.

The life history of the parasite may provide some insight into the sex-specific pattern we observe. This is a pseudaliid parasite, and although little is known about the life history of any species in this group, the high level of infection in juvenile and even neonate hosts for a number of parasite species supports the possibility of vertical transmission in milk or via the placenta (see review in [58]). For example, there is evidence for trans-placental transmission of *Halocercus lagenorhynchi* in bottlenose dolphins *(Tursiops truncatus*; [59]). At the same time, the prevalence of infection in older cetaceans of some species suggests the potential for horizontal transfer (see [58]). Analyses of *Skrjabinalius guevarai* in striped dolphins strongly suggests vertical transfer since neonates with only milk in their stomachs were found with up to 80 parasites [60]. However infection in adults suggests the possibility of horizontal transfer for this host parasite system as well. Vertical transfer may be another way in which inbreeding and a consequent higher infection

rate in female striped dolphins could impact on fitness in this species through a greater transfer of parasites to offspring

For exon-2 of the MHC Class II DQB1 locus, two alleles showed a significant association with parasite load, in this case between those with no parasites compared to those with infection at any level, and for both males and females. There are a number of other studies that also show this type of association between parasite load and class II MHC alleles, for example in association with the frequency of a DRB*1 allele in striped mice (*Rhabdomys pumilio*) infected with a gastrointestinal parasite [61]. However, the effect is not universal and likely depends on the particular relationship between the locus or loci investigated and the specific pathogen [62, 63]. The implication is a selective advantage for particular alleles in the context of specific pathogens. Consistent with this, we found that the charge properties at the pocket 4 residues in this locus varied between the two sampled populations, showing the possibility of directional selection and local adaptation as seen for two other delphinid species in an earlier study [13].

In this study we show that there is a significant female-specific association between genomic heterozygosity at 43,018 SNP loci and infestation with a parasite that can reduce lung function, even though an assessment using 23 microsatellite DNA loci showed no association. A significant g^2 value suggests that this relationship is associated with inbreeding [10]. Possible balancing selection restricted to females (given that all uninfected females were heterozygous) at a locus relevant to lactation (CIT) and likely vertical transmission of parasites from mother to offspring [60], may suggest a more direct connection to fitness if the health of offspring is affected. We have no data that could directly explain why females exposed to this pathogen are apparently more impacted by inbreeding than

males. However, female-specific effects from inbreeding have been suggested to be due to various possible factors including maternal investment [64], sex-specific gene expression [65], and sexual selection or life history [66]. Among the possible explanations, maternal investment seems most likely for mammals in general, where only females invest in the gestation and post-natal development of offspring. For the class II MHC DQB locus we find no sex-specific association, but we do find putative adaptive differences among populations (c.f. [13]), and an association between parasite load and genotype for both sexes. In this case the association may be based on directional selection. Together these data extend our understanding of the mechanisms by which genomic diversity can be associated with pathogen resistance and ultimately, fitness. If a sex bias were to be strong and consistent, it could affect sex ratios and Ne, and thereby have an impact on strategy for effective conservation of these populations.

Acknowledgments

This study was funded by the Hellenic State Scholarship Foundation through PhD grant awarded to Dr Georgios A. Gkafas. We thank Juan Antonio Balbuena for assistance with the parasite analyses. We thank the Wildlife Service of the Conselleria d'Agricultura, Medi Ambient, Canvi Climatic i Desenvolupament Rural of the Generalitat Valenciana for logistic and financial support to the Valencian cetacean strandings network. FJA and JAR are supported by Prometeo II/2015/018 project (Generalitat Valenciana, Spain). We thank Emer Rogan for samples from Ireland.

Data Accessibility

DNA sequences: Genbank accessions: MHC loci: XXXXX-XXXXX; RADseq reads: PRJNA606402. Vcf files for SNP data: Dryad: doi:10.5061/dryad.qv9s4mwb5

Table 1. Metrics from the InbreedR analyses for 43,018 SNP loci. Distribution variance of standardized Multilocus Heterozygosity [$\sigma^2(h)$], expected correlation between a fitness trait and heterozygosity (r^2_{Wh}), expected correlation between a fitness trait and inbreeding (r^2_{Wf}), Identity disequilibrium (g^2), distribution variance of heterozygosity - heterozygosity correlation coefficient (*HHC*) with confidence intervals (CI: 2.5% - 97.5%). Values in bold indicate statistical significance.

	$\sigma^2(h)$	r^2_{Wh}	r^2_{Wf}	g^2	HHC
All (68)	0.0226	0.019	0.0148	**0.030**±0.007, p=0.001 (CI: 0.016 – 0.041)	0.997±0.001 (CI:0.996 - 0.998)
Female (30)	0.0155	**0.517**	**0.387**	**0.021**±0.01, p=0.001 (CI: 0.003 – 0.040)	0.991±0.002 (CI: 0.987 – 0.996)
Male (38)	0.026	0.030	0.0236	**0.030**±0.01, p=0.001 (CI: 0.02 – 0.07)	0.994±0.001 (CI: 0.991 – 0.997)

494 **Table 2**. Genotypes of the female striped dolphin individuals of the outlier SNP within the
495 gene Citron Rho-Interacting Serine/Threonine Kinase (CIT). Colour coded for homozygous
496 (blue or yellow) and heterozygous (green). 'N' means the individual could not be scored.
497

Uninfected		Infected	
Genotype	Parasites	Genotype	Parasites
AC	0	CC	1
AC	0	AA	3
AC	0	AA	4
AC	0	AC	6
AC	0	CC	7
N	0	AA	12
AC	0	AA	16
AC	0	AA	16
AC	0	CC	21
AC	0	AA	37
N	0	AA	80
AC	0	AA	104
		AA	105
		AA	119
		CC	135
		AA	166
		AA	232
		N	256

498
499
500

Table 3. Estimated rates of Nonsynonymous (d_N) and Synonymous (d_S) substitutions for Non-Antigen-Binding Sites (Non-ABS), Pocket 4 Peptide Binding Sites (P4-PBS) and Antigen-Binding-Sites (ABS) of the exon-2 MHC Class II DQB1 locus of striped dolphin individuals. Significance was assessed using a two tailed test of the probability that d_N and d_S are different using z-test.

Position	d_N	d_S	d_N/d_S	*p; z-test value*
Non-ABS	0.026 ± 0.013	0.014 ± 0.014	1.85	0.260; z = 0.646
P4-PBS	0.268 ± 0.037	0.070 ± 0.064	3.82	0.002; z = 2.949
ABS	0.190 ± 0.054	0.009 ± 0.010	21.11	0.00058; z = 3.465

Figure legends

Figure 1: Correlation between parasite load (total number of parasites counted per individual) and RAD standardized Multilocus Heterozygosity (*sMLH*) for **a**) 30 female striped dolphin individuals and **b**) 38 male dolphins.

Figure 2: Relationship between RAD-*sMLH* mean values (±SE) against the infestation status of the striped dolphin individuals. F low (sMLH: 1.10 ± 0.013 s.e.) = females with low parasite load (≤20); F high (sMLH: 0.948 ± 0.042)= females with high load (>20); M low (sMLH: 0.966 ± 0.045)= males with low and M high (sMLH: 0.961 ± 0.027) = males with high load. Error bars show standard error of the mean.

References

1.Tsitrone A, Rousset F, David P. 2001 Heterosis, marker mutational processes and population inbreeding history. *Genetics* **159,** 1845–1859

2. David P. 1998 Heterozygosity–fitness correlations: new perspectives on old problems. *Heredity* **80,** 531–537

3. Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC. 2009 A quantitative review of heterozygosity–fitness correlations in animal populations. *Mol Ecol* **18,** 2746–2765

4. Hoffman JA, Simpson F, David P, Rijks JM, Kuiken T, Thorne MAS, Lacy RC, Dasmahapatra KK. 2014 High–throughput sequencing reveals inbreeding depression in a natural population. *PNAS* **111,** 3775–3780

5. Lozier JD. 2014 Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Mol Ecol* **23,** 788–801

6. Fischer, M.C., Rellstab, C., Leuzinger, M. et al. 2017 Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri. BMC Genomics* **18,** 69 (https://doi.org/10.1186/s12864-016-3459-7)

7. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non–Model Species. *PLoS ONE* **7,** e37135

8. Hansson B, Westerberg L. 2003 On the correlation of heterozygosity and fitness in natural populations. *Mol Ecol* **11,** 2467–2474

9. Lenz TL, Mueller B, Trillmich F, Wolf JB. 2013 Divergent allele advantage at MHC–DRB through direct and maternal genotypic effects and its consequences for allele pool composition and mating. *Proc R Soc B* **280,** 20130714 (doi: 10.1098/rspb.2013.0714)

10. Szulkin M, Bierne N, David P. 2010 Heterozygosity–fitness correlations: a time for reappraisal. *Evolution* **64,** 1202–1217

11. Bernatchez L, Landry C. 2003 MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *J Evolution Biol* **16,** 363–337

12. Eizaguirre C, Lenz TL. 2010 Major histocompatibility complex polymorphism: dynamics and consequences of parasite–mediated local adaptation in fishes. *J Fish Biol* **77,** 2023–2047

13. Vassilakos D, Natoli A, Dahlheim M, Hoelzel AR. 2009 Balancing and directional selection at exon–2 of the MHC DQB1 locus among populations of odontocete cetaceans. *Mol Biol Evol* **26,** 681–689

14. Schwensow N, Eberle M, Sommer S. 2008 Compatibility counts: MHC–associated mate choice in a wild promiscuous primate. *Proc R Soc B* **75,** 555–564

15. Juola FA, Dearborn DC. 2012 Sequence–based evidence for major histocompatibility complex disassortative mating in a colonial seabird. *Proc R Soc B* **279,** 153–162

16. Ou D, Mitchell LA, Domeier ME, Tsang AOW, Decarie D, Nepom GT, Lacroix M, Zrein M. 1996 Characterization of HLA restrictive elements of rubella virus–specific cytotoxic T cell clone: influence of HLADR4b chain residue 74 polymorphism on Ag peptide–T cell interaction. *Int Immunol* **8,** 1577–1586

17. Raga JA, Carbonell E. 1985 New data about parasites on *Stenella coeruleoalba* (Meyen, 1833) (Cetacea: Delphinidae) in the western Mediterranean Sea. *Inv Cetacea* **14**, 337–338

18. Raga JA, Abril F, Almor P. 1987 *Skrjabinalius guevaraii* Galego and Selva 1979 (Nematoda: Pseudaliidae), a lungworm parasitizing dolphins (Cetacea: Delphinidae) in the Western Mediterranean Sea. *Rivista di Parassitologia* **4**:27–32

19. Clausen B, Andersen S. 1988 Evaluation of bycatch and health status of the harbor porpoise (*Phocoena phocoena*) in Danish waters. *Dan Rev Game Biol* **13,** 1–20

20. Raga JA, Balbuena JA, Aznar FJ, Fernandez M. 1997 The impact of parasites on marine mammals: a review. *Parasitologia* **39,** 293–296

21. Pugliares K, Bogomonli A, Touhey KM, Herzig SM, Harry CT, Moore MJ. 2007 Marine Mammal Necropsy: An introductory guide for stranding responders and field biologists. *Woods Hole Oceanog Inst Tech Rept* WHOI–2007–06.

22. Gkafas GA, Exadactylos A, Rogan E, Raga JA, Reid R, Hoelzel AR. 2017 Biogeography and temporal progression during the evolution of striped dolphin population structure in European waters. *J Biogeogr* **00**:1–11.

23. Aasen E, Medrano JF. 1990 Amplification of the ZFY and ZFX genes for sex identification in human, cattle, sheep, and goats. *Nat Biotechnol* **8,** 1279–1281

24. Gilson A, Syvanen M, Levine K, Banks J. 1998 Deer gender determination by polymerase chain reaction: validation study and application to tissues, bloodstains and hair forensic samples from California. *Calif Fish Game* **84,** 159–169

25. Calzada N, Aguilar A, Lockyer C, Grau E. 1997 Patterns of grown and physical maturity in the western Mediterranean striped dolphn, *Stenella coeruleoalba* (Cetacea: Odontoceti). *Can J Zool* **75,** 632-637

26. Sambrook J, Fritsch EF, Maniatis T. 1989 Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press.

27. Siebert U. et al., 2001 Post-mortem findings in harbour porpoises (*Phocoena phocoena*) from the German North and Baltic Seas. *J Comp Pathol* **124,** 102–114

28. Lepais O, Weir JT. 2014 SimRAD: an R package for simulation–based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Mol Ecol Resour* **14,** 1314–1321

627 29. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013 Stacks: an analysis
628 tool set for population genomics. *Mol Ecol* **22,** 3124–3140

629 30, Lindblad-Toh et al. 2011 A high-resolution map of human evolutionary constraint using
630 29 mammals. *Nature* **478**, 476-482

631 31. Li H. 2012 Exploring single–sample SNP and INDEL calling with whole–genome de
632 novo assembly. *Bioinformatics* **28,** 1838–1844

633 32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
634 Durbin R, and 1000 Genome Project Data Processing Subgroup. 2009 The Sequence
635 alignment/map (SAM) format and SAMtools. *Bioinformatics* **25,** 2078-2079.

636 33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
637 Altshuler D, Gabriel S, Daly M, et al. 2010 The Genome Analysis Toolkit: a MapReduce
638 framework for analyzing next–generation DNA sequencing data. *Genome Res* **20,** 1297–1303

639 34. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007
640 TASSEL: software for association mapping of complex traits in diverse samples.
641 *Bioinformatics* **23,** 2633–2635

642 35. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker R,
643 Lunter G, Marth G, et al. 2011 The Variant Call Format and VCFtools. *Bioinformatics* **27,**
644 2156–2158

645 36. Turner SD. 2014 qqman: an R package for visualizing GWAS results using Q-Q and
646 Manhattan plots. Preprint at https://www.biorxiv.org/content/ early/2014/05/14/005165

647 37. Jombart T. 2008 adegenet: a R package for the multivariate analysis of genetic markers.
648 *Bioinformatics* **24**, 1403–1405

649 38. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL
650 2004 Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12

651 39. Stoffel MA, Esser M, Kardos M, Humble E, Nichols H, David P, Hoffman JI. 2016
652 InbreedR: An R Package for the Analysis of Inbreeding Based on Genetic Markers. *Methods*
653 *Ecol Evol* **7**,1331–1339

654 40. David P, Pujol B, Viard F, Castella V, Goudet J. 2007 Reliable selfing rate estimates from
655 imperfect population genetic data. *Mol. Ecol* **16**, 2474-2487

656 41. Balloux F, Amos W, Coulson T. 2004 Does heterozygosity estimate inbreeding in real
657 populations? *Mol Ecol* **13**, 3021-3031

658 42. Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC 2004
659 Understanding the relationship between the inbreeding coefficient and multilocus
660 heterozygosity: theoretical expectations and empirical data. *Heredity* **93**, 255

661 43. Tsuji K, Aizawa M, Sasazuki T. 1992 HLA 1991: proceedings of the 11th international
662 histocompatibility workshop and conference. Oxford: Oxford University Press.

663 44. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H,
664 Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007 Clustal W and Clustal X version 2.0.
665 *Bioinformatics* **23**, 2947–2948

666 45. Tamura K, Dudley J, Nei M, Kumar S. 2007 MEGA4: Molecular Evolutionary Genetics
667 Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599

668 46. Nei M, Gojobori T. 1986 Simple methods for estimating the numbers of synonymous and
669 non–synonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426

670 47. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. 1993
671 Three–dimensional structure of the human class II histocompatibility antigen HLA–DR1.
672 *Nature* **364**, 33–39

673 48. Stern JL, Brown JH, Jardetzky BJH, Gorga JC, Urban RG, Strominger JL, Wiley DC.
674 1994 Crystal structure of the human class II MHc protein HLA–DRB1 complexed with an
675 influenza virus peptide. *Nature 368*, 215–221

676 49. Ou D, Mitchel LA, Tingle AJ. 1998 A new categorization of HLA DR alleles on a
677 functional basis. *Hum Immunol* **59,** 665–676

678 50. Brown AA. 2015 largeQvalue: a program for calculating FDR estimates with large
679 datasets. Preprint at https://www.biorxiv.org/content/early/2015/03/18/010074

680 51. Measures LN. 2001 Lungworms of marine mammals. In: *Parasitic diseases of wild*
681 *mammals*; 2nd Edition. Samuel WM, Pybus MJ & Kocan AA Eds., The Iowa State University
682 Press. Pp. 279-300.

683 52. Richardson DS, Komdeur J, Burke T. 2004 Inbreeding in the Seychelles warbler:
684 environment–dependent maternal effects. *Evolution* **58**, 2037–2048

685 53. Jamieson IG, Roy MS, Lettink M. 2003 Sex–specific consequences of recent inbreeding
686 in an ancestrally inbred population of New Zealand takahe. *Conserv Biol* **17**, 708–716

687 54. Keller LF. 1998 Inbreeding and Its Fitness Effects in an Insular Population of Song
688 Sparrows (*Melospiza melodia*). *Evolution* **52**, 240–250

689 55. Smith TC, Fridy PC, Li Y, Basil S, Arjun S, Friesen RM, Leszyk J, Chait BT, Rout MP,
690 Luna EJ. 2013 Supervillin binding to myosin II and synergism with anillin are required for
691 cytokinesis. *MBoC* **24**, 3603–3619

692 56. Rios AC, Fu NY, Jamieson PR, Pal B, Whitehead L, Nicholas KR, Lindeman GJ,
693 Visvader JE. 2016 Essential role for a novel population of binucleated mammary epithelial
694 cells in lactation. *Nat commun* **22**, 11400

695 57. Rose AB, Beliakoff JA. 2000 Intron-mediated enhancement of gene expression
696 independent of unique intron sequences and splicing. *Plant Phys* **122**, 535-542.

697 58. Measures LN. 2001 Lungworms of marine mammals. In W. Samuel, M. Pybus, & A.
698 Kocan, eds. Parasitic Diseases of Wild Mammals. Ames, Iowa: Iowa State Press.

699 59. Dailey, M. et al., 1991 Evidence of prenatal infection in the bottlenose dolphin (Tursiops
700 truncatus) with the lungworm Halocercus lagenorhynchi (Nematoda: Pseudaliidae). *J Wildlife*
701 *Dis* **27**, 164–165

702 60. Pool R, Chandradeva N, Gkafas GA, Raga JA, Fernández M, Aznar FJ. 2019
703 Transmission and Predictors of Burden of Lungworms of the Striped Dolphin (*Stenella*
704 *coeruleoalba*) in the Western Mediterranean. *J Wildlife Dis [In print]*

705 61. Froeschke G, Sommer S. 2005 MHC Class II DRB variability and parasite load in the
706 striped mouse (*Rhabdomys pumilio*) in the Southern Kalahari. *Mol Biol Evol* **22**,1254–1259

707 62. Radwan J, Biedrzycka A, Babik W. 2009 Does reduced MHC diversity decrease viability
708 of vertebrate populations? CEE review 09–004 (SR65). Collaboration for Environmental
709 Evidence (www.environmentalevidence.org/SR65.html)

710 63. Radwan J, Kuduk K, Levy E, Lebas N, Babik W. 2014 Parasite load and MHC diversity
711 in undisturbed and agriculturally modified habitats of the ornate dragon lizard. *Mol Ecol* **23**,
712 5966–5978

713 64. Charpentier M, Setchell JM, Prugnolle F et al. 2006 Life history correlates of inbreeding
714 depression in mandrills (*Mandrillus sphinx*). *Mol Ecol* **15**, 21–28

715 65. Yun L, Agrawal AF. 2014 Variation in the strength of inbreeding depression across
716 environments: effects of stress and density dependence. *Evolution* **68**, 3599–3606

717 66. Ebel ER, Phillips PC. 2016 Intrinsic differences between males and females determine
718 sex-specific consequences of inbreeding. *BMC Evol Biol* **16**, 1

719

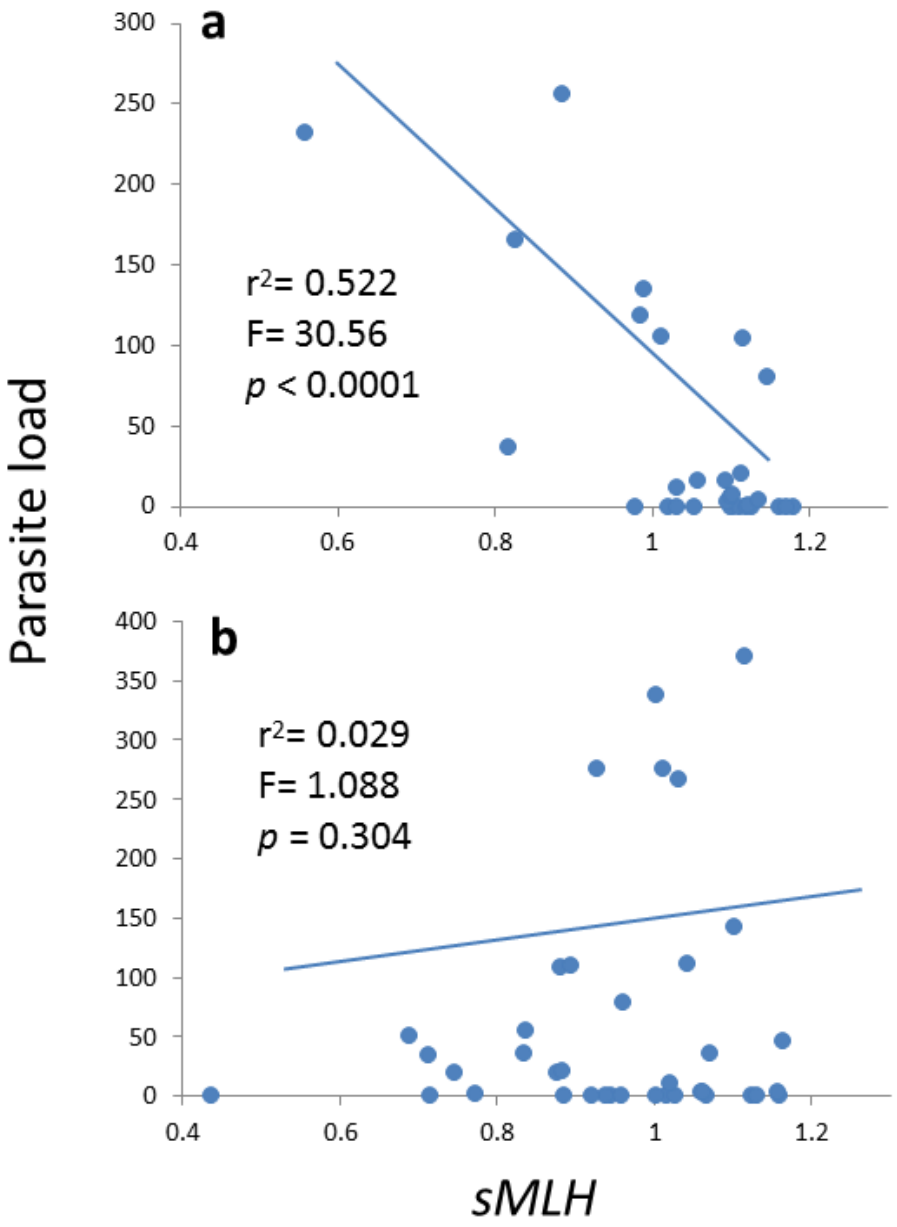
720

721

722

723

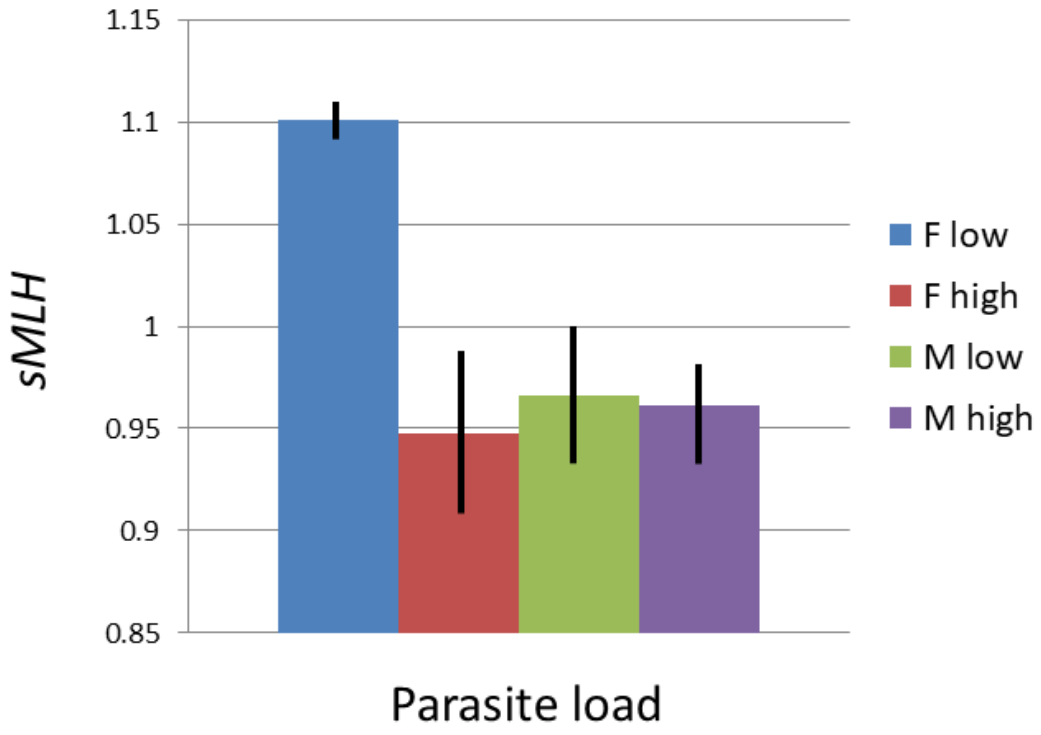
724 **Figure 1**



725

726

727 **Figure 2**



728

729

730

731